

Citation for published version:

Wood, SN 2013, 'A simple test for random effects in regression models', *Biometrika*, vol. 100, no. 4, pp. 1005-1010. <https://doi.org/10.1093/biomet/ast038>

DOI:

[10.1093/biomet/ast038](https://doi.org/10.1093/biomet/ast038)

Publication date:

2013

Document Version

Publisher's PDF, also known as Version of record

[Link to publication](#)

Publisher Rights

CC BY

© 2013 Biometrika Trust. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

A simple test for random effects in regression models

BY SIMON N. WOOD

Department of Mathematical Sciences, University of Bath, Bath BA2 7AY, U.K.

s.wood@bath.ac.uk

SUMMARY

Testing that random effects are zero is difficult, because the null hypothesis restricts the corresponding variance parameter to the edge of the feasible parameter space. In the context of generalized linear mixed models, this paper exploits the link between random effects and penalized regression to develop a simple test for a zero effect. The idea is to treat the variance components not being tested as fixed at their estimates and then to express the likelihood ratio as a readily computed quadratic form in the predicted values of the random effects. Under the null hypothesis this has the distribution of a weighted sum of squares of independent standard normal random variables. The test can be used with generalized linear mixed models, including those estimated by penalized quasilielihood.

Some key words: Generalized linear mixed model; p -value; Random effect; Variance component.

1. A STRAIGHTFORWARD TEST

Consider the linear mixed model for n independent response variables, y_i ,

$$\eta = X\beta + \sum_{j=1}^m Z_j b_j, \quad b_j \sim N(0, \psi_j \sigma^2), \quad y_i \sim N(\eta_i, \sigma^2),$$

where Z_j and X are model matrices, β is a vector of parameters and $\psi_j \sigma^2$ is a parameterized covariance matrix for the random effects, b_j . Such models are discussed in detail in [Pinheiro & Bates \(2000\)](#), for example. We wish to test $H_0: \psi_k = 0$ for some k . Exact tests require simulation and are known only for the Gaussian case, either for models with a single variance component ([Crainiceanu & Ruppert, 2004](#)) or for models with independent random effects each depending on a single variance parameter ([Wang & Chen, 2012](#)). However, practitioners routinely use models containing many variance components, with exponential family distributions other than Gaussian, and are then forced back on quite crude approximations that often lack power, as shown by [Scheipl et al. \(2008\)](#) for the Gaussian case. This paper proposes a simple simulation-free test that treats variance components, other than ψ_k , as fixed at their estimated values, but extends to the case of generalized linear mixed models with multiple variance components.

Restricted maximum likelihood or maximum likelihood can be used to estimate the ψ_j . Given these estimates, the predicted random effects \hat{b}_j and the maximum likelihood estimates $\hat{\beta}$ can be found by minimization of

$$\left\| y - X\beta - \sum_{j=1}^m Z_j b_j \right\|^2 + \sum_{j=1}^m b_j^T \hat{\psi}_j^{-1} b_j$$

with respect to the b_j and β , where $\hat{\psi}_j^{-1}$ is the inverse, or sometimes the Moore–Penrose pseudo-inverse, of $\hat{\psi}_j$. If $\hat{B} = (\hat{\beta}, \hat{b}_1, \hat{b}_2, \dots)$ then the solution to this minimization can be written $\hat{B} = Py$, where the matrix P is given in the next paragraph. Let P_j denote the rows of P such that $\hat{b}_j = P_j y$. Under the null hypothesis $y \sim N(X\beta, \Sigma_{-k})$, where $\Sigma_{-k} = (I + \sum_{j \neq k} Z_j \psi_j Z_j^T) \sigma^2$, so $\hat{b}_k \sim N(0, P_k \Sigma_{-k} P_k^T)$, since $E(\hat{b}_k) = 0$, as demonstrated next.

Let $Z = (Z_1 : Z_2 : \dots)$ with corresponding combined random effects vector b . To prove that $E(\hat{b}) = 0$, note that $\hat{\mathcal{B}}$ is the minimizer of the augmented residual sum of squares

$$\left\| \begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} X & Z \\ 0 & B \end{pmatrix} \begin{pmatrix} \beta \\ b \end{pmatrix} \right\|^2$$

with respect to β and b . The matrix B is a matrix square root of $\sum_j \hat{\psi}_j^-$. So we have

$$\tilde{X} = \begin{pmatrix} X & Z \\ 0 & B \end{pmatrix}, \quad \begin{pmatrix} \hat{\beta} \\ \hat{b} \end{pmatrix} = (\tilde{X}^\top \tilde{X})^{-1} \tilde{X}^\top \begin{pmatrix} y \\ 0 \end{pmatrix}, \quad E \begin{pmatrix} y \\ 0 \end{pmatrix} = \tilde{X} \begin{pmatrix} \beta \\ b \end{pmatrix}$$

and hence $E(\hat{b}) = 0$. The matrix P is given by the first n columns of $(\tilde{X}^\top \tilde{X})^{-1} \tilde{X}^\top$.

Now consider a testing procedure. The proposal here is that σ and ψ_j for $j \neq k$ should be treated as fixed at their estimated values. Then the test statistic is based on the log of the ratio of the restricted likelihood under the null hypothesis, $H_0 : \psi_k = 0$, and the alternative hypothesis that $H_1 : \psi_k = \hat{\psi}_k$. Under H_1 the restricted loglikelihood can be expressed in terms of $\hat{\beta}$ and \hat{b}_j as

$$\hat{l}'_1 = -\frac{1}{2\hat{\sigma}^2} \left\| y - X\hat{\beta} - \sum_{j=1}^m Z_j \hat{b}_j \right\|^2 - \frac{1}{2\hat{\sigma}^2} \sum_{j=1}^m \hat{b}_j^\top \hat{\psi}_j^- \hat{b}_j + c_1 = \frac{\hat{l}_1}{\hat{\sigma}^2} + c_1,$$

while under H_0 it is

$$\hat{l}'_0 = -\frac{1}{2\hat{\sigma}^2} \left\| y - X\tilde{\beta} - \sum_{j \neq k} Z_j \tilde{b}_j \right\|^2 - \frac{1}{2\hat{\sigma}^2} \sum_{j \neq k} \tilde{b}_j^\top \hat{\psi}_j^- \tilde{b}_j + c_0 = \frac{\hat{l}_0}{\hat{\sigma}^2} + c_0.$$

These expressions can be derived by integrating \mathcal{B} out of the joint density of y and the random effects via a Laplace approximation, which is exact for a Gaussian linear mixed model. See Wood (2011) for further details. The constants c_0 and c_1 are irrelevant, given fixed ψ_j and σ . The vectors $\hat{\beta}$ and \hat{b}_j are the maximizers of l'_1 , while $\tilde{\beta}$ and \tilde{b}_j are the maximizers of l'_0 . The test statistic is $\mathcal{W} = 2(\hat{l}_1 - \hat{l}_0)$, which can be expressed as a quadratic form in \hat{b}_k , as shown next, by exploiting standard numerical linear algebra (e.g., Wood, 2006, Appendix A).

Without loss of generality, assume that $k = m$, the largest value of j : the Z_j can always be reordered to ensure that this is so. Let $p = \dim(\mathcal{B})$ and $p_k = \dim(b_k)$. Consider the QR decomposition of the augmented model matrix, i.e.,

$$\tilde{X} = \begin{pmatrix} X & Z_{-k} & Z_k \\ 0 & B_{-k} & 0 \\ 0 & 0 & B_k \end{pmatrix} = Q \begin{pmatrix} R \\ 0 \end{pmatrix}$$

where Z_{-k} is Z with the columns of Z_k omitted, B_{-k} is a square root of $\sum_{j \neq k} \hat{\psi}_j^-$ and B_k a square root of $\hat{\psi}_k^-$. Now $\hat{\beta}$ and the \hat{b}_j are the minimizers of $\|f - R\mathcal{B}\|^2$, where f^\top is the first p elements of $(y^\top, 0)Q$ and we define r to be the remaining n elements. Hence $\hat{\mathcal{B}} = R^{-1}f$ and $2\hat{l}_1 = -\|r\|^2$. Now partition

$$R = \begin{pmatrix} R_0 & R_1 \\ 0 & \tilde{R} \end{pmatrix}$$

where R_0 is the upper left $p - p_k \times p - p_k$ block of R . Letting f_0 denote the first $p - p_k$ rows of f and f_1 the remaining p_k rows, then $\tilde{\beta}$ and the \tilde{b}_j are the minimizers of

$$\|f_0 - R_0 \mathcal{B}_{-k}\|^2$$

where \mathcal{B}_{-k} is \mathcal{B} with b_k omitted, while $2\hat{l}_0 = -\|r\|^2 - \|f_1\|^2$. Hence $2(\hat{l}_1 - \hat{l}_0) = \|f_1\|^2$. By the upper triangular structure of R , $f_1 = \tilde{R}\hat{b}_k$, so $2(\hat{l}_1 - \hat{l}_0) = \hat{b}_k^\top \tilde{R}^\top \tilde{R} \hat{b}_k$. Now find a matrix square root, C , of the covariance matrix of \hat{b}_k under H_0 , so that $C^\top C = P_k \Sigma_{-k} P_k^\top$. If z_k is a p_k vector of independent $N(0, 1)$ random variables, it follows that under H_0 ,

$$\mathcal{W} = \hat{b}_k^\top \tilde{R}^\top \tilde{R} \hat{b}_k \sim z_k^\top C \tilde{R}^\top \tilde{R} C^\top z_k = z_k^\top U \Lambda U^\top z_k \sim \sum_{i=1}^{p_k} \lambda_i \chi_{1i}^2, \quad (1)$$

where $U \Lambda U^\top$ is the spectral decomposition of $C \tilde{R}^\top \tilde{R} C^\top$, so that U is an orthogonal matrix while Λ is a diagonal matrix of eigenvalues, λ_i . The cumulative distribution function of such a weighted sum of χ_1^2 random variables can be computed by the method of Davies (1980), or approximated using Liu et al. (2009). If ψ_j and σ were known, then this test would be exact and is simply a likelihood ratio test. Result (1) can also be obtained by first integrating out the random effects, other than the k th, so that we start with

$$\tilde{X} = \begin{pmatrix} \hat{\sigma} \hat{\Sigma}_{-k}^{-1/2} X & \hat{\sigma} \hat{\Sigma}_{-k}^{-1/2} Z_k \\ 0 & B_k \end{pmatrix},$$

where $\hat{\Sigma}_{-k}^{-1/2}$ is the inverse of a matrix square root of $\hat{\Sigma}_{-k}$.

When σ^2 has been estimated, slightly improved p -values can be obtained by quadrature evaluation of the one-dimensional integral required to obtain $\text{pr}(\sum_i \lambda_i \chi_{1i}^2 > \mathcal{W}_{\text{obs}} \chi_\kappa^2 / \kappa)$ where κ is the residual degrees of freedom used in the σ^2 estimation, and \mathcal{W}_{obs} is the observed value of \mathcal{W} . This has negligible computational cost given the ability to rapidly compute the cumulative distribution function of $\sum_i \lambda_i \chi_{1i}^2$ provided by Davies (1980) or Liu et al. (2009).

Computationally the approach is most efficient if X and the Z_j are not used directly. Rather the QR decomposition $\mathcal{QR} = (X, Z_1, Z_2, \dots)$ is performed first, or simply retained from the original model fit, and X and the Z_j are then replaced by the corresponding columns of \mathcal{R} , in all expressions. This minimizes the cost of reordering and the subsequent QR decomposition to find \tilde{R} , and of computing the Σ_{-k} . Further computational efficiency can be obtained, albeit at some cost in numerical stability, by noting that R is also the Cholesky factor of $\tilde{X}^\top \tilde{X}$. The computational limit is then set either by the size of matrix for which Cholesky decomposition is practical, or by the size of symmetric eigenvalue problem that is practical.

The extension to generalized linear mixed models is routine. Let the y_i be conditionally independent random variables from an exponential family distribution with mean μ_i determined by $g(\mu_i) = \eta_i$, scale parameter ϕ and variance function $V(\mu)$. The function g is a known link function. The Laplace approximate restricted loglikelihood under H_1 is

$$\hat{l}_1 \simeq \log f(y | \hat{\mathcal{B}}) - \frac{1}{2\hat{\phi}} \sum_{j=1}^m \hat{b}_j^\top \hat{\psi}_j^- \hat{b}_j + c'_1$$

where $f(y | \mathcal{B})$ is the conditional density of y given \mathcal{B} and c'_1 is a constant that is ignorable when treating the variance parameters as fixed. Under H_0 the equivalent is

$$\hat{l}_0 \simeq \log f(y | \tilde{\mathcal{B}}) - \frac{1}{2\hat{\phi}} \sum_{j \neq k}^m \tilde{b}_j^\top \hat{\psi}_j^- \tilde{b}_j + c'_0.$$

Assuming fixed dimension for \mathcal{B} the ratio of these approximations to the truth is $1 + O(n^{-1})$, and further details are given in Wood (2011).

With $z_i = g'(\hat{\mu}_i)(y - \hat{\mu}_i) + \hat{\eta}_i$ and $W_i^2 = g'(\hat{\mu}_i)^{-2} V(\hat{\mu}_i)^{-1}$, a standard quadratic approximation of $\log f(y | \mathcal{B})$ about $\hat{\mathcal{B}}$ yields

$$\hat{l}_1 \simeq -\frac{1}{2\hat{\phi}} \left\| W \left(z - X\hat{\beta} - \sum_{j=1}^m Z_j \hat{b}_j \right) \right\|^2 - \frac{1}{2\hat{\phi}} \sum_{j=1}^m \hat{b}_j^\top \hat{\psi}_j^- \hat{b}_j + c''_1.$$

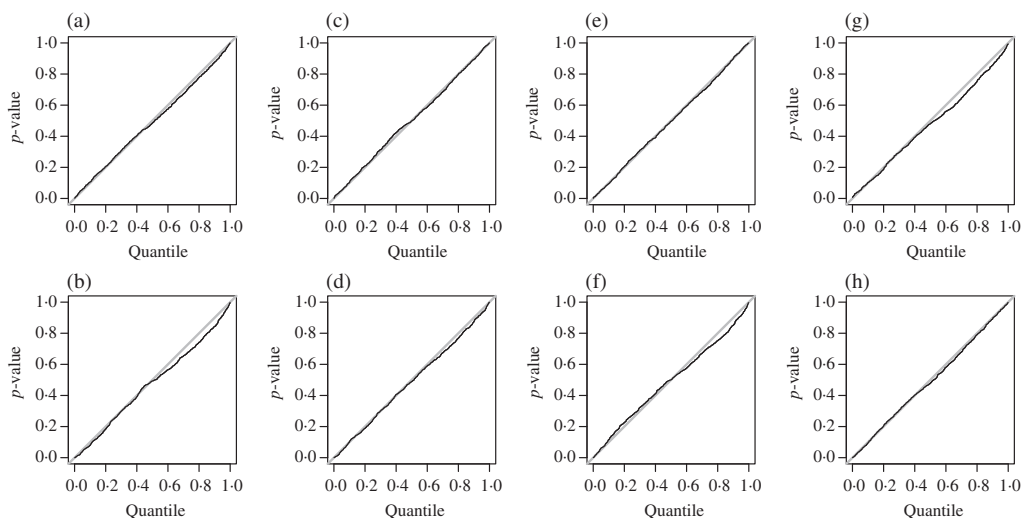


Fig. 1. Null distribution of p -values. Quantile-quantile plots for the null distribution of p -values of the new method overlaid over a grey reference line. Panels (a)–(f) are for Gaussian responses following § 3.1.1 of Scheipl et al. (2008). Sample sizes are 30 (a, b), 100 (c, d) and 500 (e, f). Group sizes from (a) to (f) are 3, 10, 5, 20, 5, 100. Panels (g) and (h) are for binary responses. Sample sizes are 100 for (g) and 400 for (h) with a group size of 20.

Under H_0 , the same quadratic approximation remains valid for \hat{l}'_0 so that

$$\hat{l}'_0 \simeq -\frac{1}{2\hat{\phi}} \left\| W \left(z - X\tilde{\beta} - \sum_{j \neq k}^m Z_j \tilde{b}_j \right) \right\|^2 - \frac{1}{2\hat{\phi}} \sum_{j \neq k}^m \tilde{b}_j^\top \hat{\psi}_j^- \tilde{b}_j + c''_0.$$

Hence, given large sample normality of the \hat{b}_k (e.g., Wood, 2006, § 4.8), the test proceeds exactly as in the linear case, but with WX replacing X and WZ_j replacing Z_j in all computations.

2. SIMULATION TESTING

The simulation study of Scheipl et al. (2008) was repeated, to compare the test proposed here with a test based on simulation of the correct distribution of the likelihood ratio statistic, and with the best test found in Scheipl. The latter is an approximation proposed by Greven et al. (2008) based on the exact test of Crainiceanu & Ruppert (2004) and Crainiceanu et al. (2005) and implemented in R (R Development Core Team, 2013) package RLRsim. To test the k th random effect the approach replaces the response variable by $y - \sum_{j \neq k} Z_j \hat{b}_j$ where the \hat{b}_j are predicted random effects. Crainiceanu & Ruppert (2004) is then applied to test the single remaining effect for equality to zero. This approach is difficult to extend beyond Gaussian response distributions, while in the Gaussian case, the key underlying difference to the method proposed here is neglect of the variability associated with the predictions, \hat{b}_j .

The full study results are in the online Supplementary Material, and include cases with multiple variance components. Figures 1(a)–(f) and 2(a)–(f) show the results for the single variance component case in which Crainiceanu & Ruppert (2004) is exact. The model is $y_i = \alpha + b_j + \epsilon_i$ where y_i is assumed to belong to group j , b_j is a Gaussian random effect, and ϵ_i a Gaussian residual error term. The single variance component is the variance of the b_j . Power curves are shown against a standardized effect size, but were obtained by varying the standard deviation of the b_j from zero up to 4.5, 1.8 and 0.75, respectively, for the sample sizes 30, 100 and 500. In all cases in the full study the test proposed here gives power curves very close to the Scheipl recommended method, and to the curve resulting from direct simulation, but since

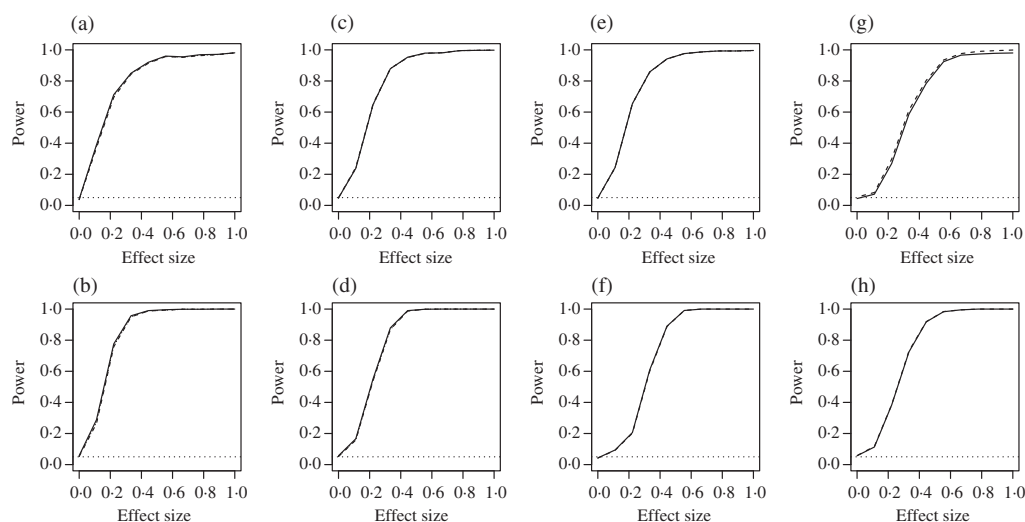


Fig. 2. Power comparisons with Crainiceanu & Ruppert (2004) and with direct simulation. Data and models for each panel correspond to those given for the panels in Fig. 1. Power curves, at the 5% level, are plotted as continuous for the new method, dotted for Crainiceanu & Ruppert (2004) and dashed for direct simulation. The curves are often so similar as to be indistinguishable. Effect strength was controlled by increasing the random effect variance from zero, but is shown on a standardized scale. The horizontal dotted line denotes the 5% level. Panels (g) and (h) compare the proposed method with a method using exact simulation of the null distribution of the likelihood ratio statistic for logistic regression with binary data, since Crainiceanu & Ruppert (2004) does not apply in this case.

variance parameter estimator uncertainty has been neglected there is a slight tendency for p -values to be too low. For cases in which variance components are poorly identified this can give elevated Type I error rates, and the online Supplementary Material also includes a simulation designed to illustrate this.

Figures 1(g)–(h) and 2(g)–(h) illustrate the test behaviour for a logistic regression model fitted to binary data simulated from a generalized linear mixed model with logit link and linear predictor $\eta_i = f(z_i) + b_j$ where f is a smooth function of predictor variable z , while b_j is a 20 level Gaussian random effect, the standard deviation of which was varied upwards from zero to 5 and 2.5 respectively for sample sizes 100 and 400, to produce 5% power curves. Beyond the Gaussian case there is no exact test known, so the new test was compared to full simulation for the null distribution of the Laplace approximate restricted likelihood ratio. The proposed method shows some performance degradation at sample size 100, attributable to deterioration of the normality approximation for the \hat{b}_j at such sample sizes.

3. DISCUSSION

The test suggested here provides a way of testing variance components in generalized linear mixed models, including models estimated by penalized quasiliquelihood, and has the practical advantage of being computable without multiple model fits or simulation of the null distribution. Its main drawback is that it treats the variance components as fixed at their estimates. Although practitioners routinely treat the variance components in this way for inference about the fixed effects, it would clearly be desirable to eliminate this assumption. However, until this is possible other than by direct simulation, the proposal is at least a readily computed, well-founded test, where the assumptions are explicit.

The test is likely to show reduced reliability in three circumstances: firstly, if the variance parameter estimates are very poor, as is the case for penalized quasiliquelihood estimates from binary data; secondly, when normality of the \hat{b}_k is a poor approximation, for example for small binary datasets where some \hat{b}_k may not even be well defined; thirdly, if ψ_k is highly uncertain and highly correlated with other ψ_j , as is the case when covariates are highly correlated and the sample size is relatively small. The Supplementary Material

provides an example of this. In such cases simulation should be used. Another limitation, shared with previous approaches, is that the proposed test is not suitable for zero effect testing of only some variance components from a set of effects modelled as correlated. For example, in a model with correlated slope and intercept effects, it is not possible to directly test for zero variance of the random slopes. However, in practice it is usual to test for independence of effects before testing for no effect, and a standard generalized likelihood ratio test is unproblematic for the independence test: the Supplementary Material provides an example.

Given that the variance component estimates have known asymptotic sampling distributions, it might be interesting to investigate p -values computed by model averaging in order to allow for the variance component uncertainty that is neglected here. Another issue is the utility of these tests for very large random effects models estimable only by sparse matrix methods. Apart from the final eigendecomposition, the test computations require only Cholesky decomposition, for which reliable sparse algorithms exist, but the need to pivot to maintain sparseness presents a substantial challenge, since pivoting spoils the ordering of R on which the test relies. The eigendecomposition is also potentially challenging, but could probably be achieved efficiently by Lanczos iteration, if the pivoting issue could be resolved.

ACKNOWLEDGEMENT

I am grateful to Martijn Wieling and Bryan Wood for, in different ways, initiating this work, and to a referee, an associate editor and the editor for valuable comments on earlier drafts. The work was funded by the UK Engineering and Physical Sciences Research Council and is part of the research programme of the UK National Centre for Statistical Ecology.

SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes the full results of repeating the simulation study design of Scheipl et al. (2008). Further simulation tests are presented for the cases of binary data, highly correlated data and dependent random effects. R code providing an illustration of the method implementation is also provided.

REFERENCES

- CRAINICEANU, C. & RUPPERT, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *J. R. Statist. Soc. B* **66**, 165–85.
- CRAINICEANU, C., RUPPERT, D., CLAESKENS, G. & WAND, M. P. (2005). Exact likelihood ratio tests for penalised splines. *Biometrika* **92**, 91–103.
- DAVIES, R. B. (1980). Algorithm AS 155: The distribution of a linear combination of χ^2 random variables. *J. R. Statist. Soc. C* **29**, 323–33.
- GREVEN, S., CRAINICEANU, C., KÜCHENHOFF, H. & PETERS, A. (2008). Restricted likelihood ratio testing for zero variance components in linear mixed models. *J. Comp. Graph. Statist.* **17**, 870–91.
- LIU, H., TANG, Y. & ZHANG, H. H. (2009). A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Comp. Statist. Data Anal.* **53**, 853–6.
- PINHEIRO, J. & BATES, D. (2000). *Mixed-Effects Models in S and S-PLUS*. New York: Springer.
- R DEVELOPMENT CORE TEAM (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org>.
- SCHEIPL, F., GREVEN, S. & KÜCHENHOFF, H. (2008). Size and power of tests for a zero random effect variance or polynomial regression in additive and linear mixed models. *Comp. Statist. Data Anal.* **52**, 3283–99.
- WANG, Y. & CHEN, H. (2012). On testing an unspecified function through a linear mixed effects model with multiple variance components. *Biometrics* **68**, 1113–25.
- WOOD, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton: CRC/Chapman & Hall.
- WOOD, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Statist. Soc. B* **73**, 3–36.

[Received July 2012. Revised June 2013]